Aix*Marseille université

Lecture:
Representation and approximation of structured data

Part 2

Kai Schneider

Master 1 Data Science, 2023/2024

Ref: from Vetterli & Goyal, 'Teaching Signal Processing with Geometry'

Textbook: M. Vetterli, J. Kovacevic and V. Goyal, Foundations of Signal Processing, Cambridge University Press, 2014.
http://fourierandwavelets.org/

Gram-Schmidt orthogonalization

Motivation: construction of an orthonormal basis from a set of linearly independent vectors.

Let $\mathcal{K} = \{0, 1, ..., N-1\}$

The goal is to find an orthonormal set $\{\varphi_k\}_{k\in\mathcal{K}}$ with

$$\overline{\text{span}}(\{\varphi_k\}_{k\in\mathcal{K}}) \;=\; \overline{\text{span}}(\{x_k\}_{k\in\mathcal{K}}).$$

Thus, when $\{x_k\}_{k\in\mathcal{K}}$ is a basis for H, the constructed set $\{\varphi_k\}_{k\in\mathcal{K}}$ is an orthonormal basis for H; otherwise, $\{\varphi_k\}_{k\in\mathcal{K}}$ is an orthonormal basis for the smaller space $\overline{\text{span}}(\{x_k\}_{k\in\mathcal{K}})$, which is itself a Hilbert space.

For φ_0 to have unit norm, it is natural to choose

$$\varphi_0 \;=\; x_0/\|x_0\|,$$

and for φ_1 we obtain

$$\varphi_1 \;=\; \frac{x_1 - \langle x_1, \varphi_0\rangle\varphi_0}{\|x_1 - \langle x_1, \varphi_0\rangle\varphi_0\|}$$

In general φ_k is determined by normlizing the residual of x_k orthogonally projected to $\text{span}(\{\varphi_0, \varphi_1, \ldots, \varphi_{k-1}\})$

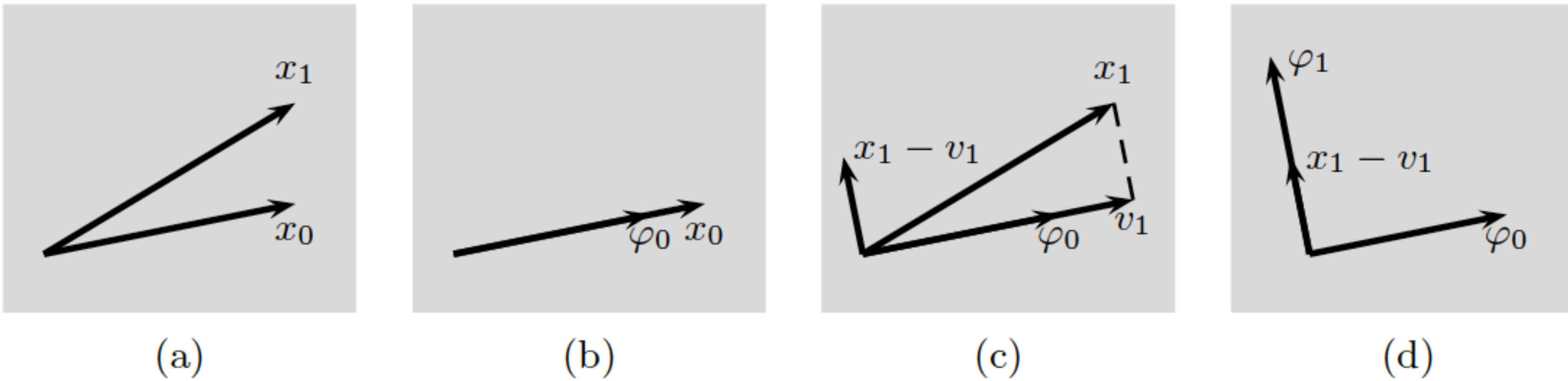

Figure 2.21 Illustration of Gram–Schmidt orthogonalization. (a) Input vectors (x_0, x_1). (b) The first output vector φ_0 is a normalized version of x_0. (c) The projection of x_1 onto the subspace spanned by φ_0 is subtracted from x_1 to obtain a residual $x_1 - v_1$. (d) The second output vector φ_1 is a normalized version of the residual.

Gram-Schmidt algorithm

Gram–Schmidt orthogonalization

Input: An ordered sequence of linearly independent vectors $(x_k)_{k \in \mathcal{K}}$

Output: Orthonormal vectors $\{\varphi_k\}_{k \in \mathcal{K}}$, with $\text{span}(\{\varphi_k\}) = \text{span}(\{x_k\})$

$\{\varphi_k\} = \textbf{GramSchmidt}(\{x_k\})$

$\varphi_0 \;=\; x_0/\|x_0\|$

$k = 1$

while $k < |\mathcal{K}|$ **do**

  project $v_k \;=\; \sum_{i=0}^{k-1} \langle x_k, \varphi_i \rangle \varphi_i$

  normalize $\varphi_k \;=\; (x_k - v_k)/\|x_k - v_k\|$

  increment k

end while

return $\{\varphi_k\}$

Example: Gram-Schmidt algorithm (order plays a role)

(i) Let $x_0 = \begin{bmatrix}1 & 1 & 0\end{bmatrix}^\top$, $x_1 = \begin{bmatrix}0 & 1 & 1\end{bmatrix}^\top$, and $x_2 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\top$. These are linearly independent, and following the steps in Table 2.1 first yields $\varphi_0 = (1/\sqrt{2})\begin{bmatrix}1 & 1 & 0\end{bmatrix}^\top$, then $v_1 = \frac{1}{2}\begin{bmatrix}1 & 1 & 0\end{bmatrix}^\top$, then $x_1 - v_1 = \frac{1}{2}\begin{bmatrix}-1 & 1 & 2\end{bmatrix}^\top$, and $\varphi_1 = (1/\sqrt{6})\begin{bmatrix}-1 & 1 & 2\end{bmatrix}^\top$. For the final basis vector, $v_2 = \frac{1}{3}\begin{bmatrix}2 & 4 & 2\end{bmatrix}^\top$, $x_2 - v_2 = \frac{1}{3}\begin{bmatrix}1 & -1 & 1\end{bmatrix}^\top$, and $\varphi_2 = (1/\sqrt{3})\begin{bmatrix}1 & -1 & 1\end{bmatrix}^\top$.

(ii)

Starting with $x_0 = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\top$, $x_1 = \begin{bmatrix}1 & 1 & 0\end{bmatrix}^\top$, and $x_2 = \begin{bmatrix}0 & 1 & 1\end{bmatrix}^\top$, the same set of vectors as in (i), but in a different order, yields

$$\varphi_0 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \varphi_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}, \quad \text{and} \quad \varphi_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix}.$$

There is no obvious relationship between this orthonormal basis and the one found in part (i).

Biorthogonal pairs of bases

In some cases the nonorthogonal basis can be easier to store and to compute with, for example

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

A basis does not have to be orthogonal to provide unique expansions. However in this case we cannot ask for a single set of vectors for the analysis and for the synthesis of vectors.

This leads to the concept of a biorthogonal pair of bases, or dual bases.

Definition 2.43 (Biorthogonal pair of bases) The sets of vectors $\Phi = \{\varphi_k\}_{k\in\mathcal{K}} \subset H$ and $\widetilde{\Phi} = \{\widetilde{\varphi}_k\}_{k\in\mathcal{K}} \subset H$, where $\mathcal{K}$ is finite or countably infinite, are called a *biorthogonal pair of bases* for a Hilbert space H when

(i) each is a *basis* for H; and

(ii) they are *biorthogonal*, meaning that

$$\langle \varphi_i, \widetilde{\varphi}_k \rangle = \delta_{i-k} \qquad \text{for every } i,\, k \in \mathcal{K}. \tag{2.111}$$

A biorthogonal pair of bases thus yields four operators (analysis and sythesis):

$$\Phi,\ \Phi^*,\ \widetilde{\Phi},\ \text{and}\ \widetilde{\Phi}^*$$

Example 2.39 (Biorthogonal pair of bases in finite dimensions) The sets

$$\varphi_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},\ \varphi_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ \varphi_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \widetilde{\varphi}_0 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix},\ \widetilde{\varphi}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},\ \widetilde{\varphi}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

are a biorthogonal pair of bases for $\mathbb{C}^3$, as can be verified by direct computation.

Expansion and inner product computation

With a biorthogonal pair of bases, expansion coefficients with respect to one basis are computed using the other basis.

THEOREM 2.44 (BIORTHOGONAL BASIS EXPANSIONS) Let $\Phi = \{\varphi_k\}_{k\in\mathcal{K}}$ and $\widetilde{\Phi} = \{\widetilde{\varphi}_k\}_{k\in\mathcal{K}}$ be a biorthogonal pair of bases for a Hilbert space H. The unique expansion with respect to the basis Φ of any x in H has expansion coefficients

$$\alpha_k = \langle x, \widetilde{\varphi}_k\rangle \qquad \text{for } k \in \mathcal{K}, \quad \text{or,} \tag{2.113a}$$

$$\alpha = \widetilde{\Phi}^* x. \tag{2.113b}$$

Synthesis with these coefficients yields

$$x = \sum_{k\in\mathcal{K}} \langle x, \widetilde{\varphi}_k\rangle \varphi_k \tag{2.114a}$$

$$= \Phi\alpha = \Phi\widetilde{\Phi}^* x. \tag{2.114b}$$

Note that we have:

$$\Phi\widetilde{\Phi}^* = I \quad \text{as well as} \quad \widetilde{\Phi}\Phi^* = I \qquad \text{on} \quad H.$$

THEOREM 2.45 (PARSEVAL EQUALITIES FOR BIORTHOGONAL PAIRS OF BASES) Let $\Phi = \{\varphi_k\}_{k\in\mathcal{K}}$ and $\widetilde{\Phi} = \{\widetilde{\varphi}_k\}_{k\in\mathcal{K}}$ be a biorthogonal pair of bases for a Hilbert space H. Expansion with respect to the bases Φ and $\widetilde{\Phi}$ with coefficients (2.113) and (2.115) satisfies

$$\|x\|^2 = \sum_{k\in\mathcal{K}} \langle x, \varphi_k\rangle\langle x, \widetilde{\varphi}_k\rangle^* \tag{2.118a}$$

$$= \langle \Phi^* x, \widetilde{\Phi}^* x\rangle = \langle \widetilde{\alpha}, \alpha\rangle. \tag{2.118b}$$

More generally,

$$\langle x, y\rangle = \sum_{k\in\mathcal{K}} \langle x, \varphi_k\rangle\langle y, \widetilde{\varphi}_k\rangle^* \tag{2.119a}$$

$$= \langle \Phi^* x, \widetilde{\Phi}^* y\rangle = \langle \widetilde{\alpha}, \beta\rangle. \tag{2.119b}$$

Best approximation and the normal equations

According to the projection theorem, the best approximation of a vector x is given by the orthogonal projection onto the subspace S. This methodology can be genelized using bases.

THEOREM 2.48 (NORMAL EQUATIONS) Given a vector x and a Riesz basis $\{\varphi_k\}_{k\in\mathcal{I}}$ for a closed subspace S in a separable Hilbert space H, the vector closest to x in S is

$$\widehat{x} = \sum_{k\in\mathcal{I}} \beta_k \varphi_k \tag{2.137a}$$

$$= \Phi\beta, \tag{2.137b}$$

where β is the unique solution to the system of equations

$$\sum_{k\in\mathcal{I}} \beta_k \langle \varphi_k, \varphi_i \rangle = \langle x, \varphi_i \rangle \qquad \text{for every } i \in \mathcal{I}, \quad \text{or,} \tag{2.138a}$$

$$\Phi^*\Phi\beta = \Phi^* x. \tag{2.138b}$$

Normal equations!

Example: normal equations (I)

EXAMPLE 2.46 (NORMAL EQUATIONS IN $\mathbb{R}^3$) Let $\varphi_0 = \begin{bmatrix}1 & 1 & 0\end{bmatrix}^\top$ and $\varphi_1 = \begin{bmatrix}0 & 1 & 1\end{bmatrix}^\top$. Given a vector $x = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\top$, according to Theorem 2.48, the vector in span$(\{\varphi_0, \varphi_1\})$ closest to x is

$$\widehat{x} = \beta_0\varphi_0 + \beta_1\varphi_1,$$

with β the unique solution to (2.138b), which simplifies to

$$\begin{bmatrix}2 & 1 \\ 1 & 2\end{bmatrix}\begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix} = \begin{bmatrix}2 \\ 2\end{bmatrix}.$$

Solving the system yields $\beta_0 = \beta_1 = \frac{2}{3}$, leading to

$$\widehat{x} = \frac{2}{3}(\varphi_0 + \varphi_1) = \begin{bmatrix}\frac{2}{3} \\ \frac{4}{3} \\ \frac{2}{3}\end{bmatrix}.$$

Example: normal equations (II)

We can easily check that the residual $x - \widehat{x}$ is orthogonal to span($\{\varphi_0, \varphi_1\}$):

$$x - \widehat{x} = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \quad \perp \quad \alpha_0\varphi_0 + \alpha_1\varphi_1 = \begin{bmatrix} \alpha_0 \\ \alpha_0 + \alpha_1 \\ \alpha_1 \end{bmatrix}.$$

Now let $\varphi_2 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^\top$. The vector in span($\{\varphi_0, \varphi_1, \varphi_2\}$) closest to x is

$$\widehat{x} = \beta_0\varphi_0 + \beta_1\varphi_1 + \beta_2\varphi_2,$$

where β satisfies

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}.$$

The solutions for β are not unique, but all solutions yield $\widehat{x} = \begin{bmatrix} \frac{2}{3} & \frac{4}{3} & \frac{2}{3} \end{bmatrix}^\top$ as before. This is as expected, since span($\{\varphi_0, \varphi_1, \varphi_2\}$) = span($\{\varphi_0, \varphi_1\}$).

Matrix representations of vectors and linear operators

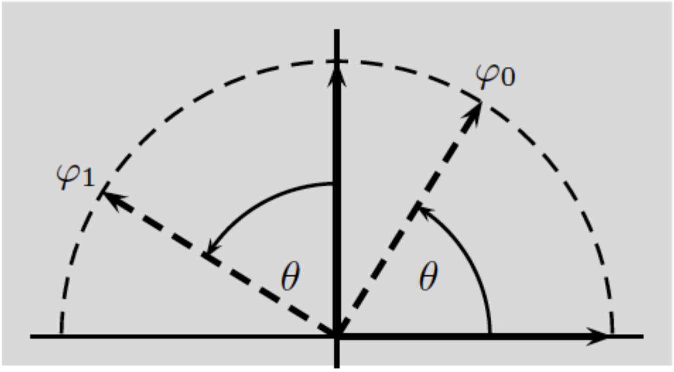

Figure 2.27 An orthonormal basis in $\mathbb{R}^2$ (dashed lines) generated by rotation of the standard basis (solid lines).

Example 2.53 (Change of basis by rotation) Let $\{\varphi_0, \varphi_1\}$ be the basis for $\mathbb{R}^2$ shown in Figure 2.27,

$$\varphi_0 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \qquad \varphi_1 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

Let $\{\psi_0, \psi_1\}$ be the standard basis for $\mathbb{R}^2$. The change of basis matrix from Φ to Ψ is

$$C_{\Phi,\Psi} = \Psi^*\Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Consider the vector in $\mathbb{R}^2$ that has representation $\alpha = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ *with respect to* Φ (not with respect to the standard basis). This means that the vector is

$$x = 1 \cdot \varphi_0 + 0 \cdot \varphi_1 = \varphi_0 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix},$$

where the final expression is with respect to the standard basis Ψ. This agrees with the result of the multiplication $C_{\Phi,\Psi}\alpha$.

Multiplying by $C_{\Phi,\Psi}$ is a counterclockwise rotation by angle θ. This agrees with the fact that the basis Φ is the standard basis Ψ rotated counterclockwise by θ.

Matrices: eigenvalues – singular values (I)

Eigenvalues, eigenvectors, and spectral decomposition A number λ and a nonzero vector v are called an *eigenvalue* and an *eigenvector* of a square matrix A (they are also known as an *eigenpair*) when

$$Av = \lambda v, \tag{2.226}$$

as seen for general linear operators in (2.58). The eigenvalues are the roots of the *characteristic polynomial* $\det(xI - A)$. When all eigenvalues of A are real, $\lambda_{\max}(A)$ denotes the largest eigenvalue and $\lambda_{\min}(A)$ the smallest eigenvalue. When the eigenvalues are real, it is conventional to list them in nonincreasing order, $\lambda_0(A) \geq \lambda_1(A) \geq \cdots \geq \lambda_{N-1}(A)$.

If an $N \times N$ matrix A has N linearly independent eigenvectors, then it can be written as

$$A = V\Lambda V^{-1}, \tag{2.227a}$$

Matrices: eigenvalues – singular values (II)

Singular value decomposition *Singular value decomposition (SVD)* provides a diagonalization that applies to any rectangular or square matrix. An $M \times N$ real or complex matrix A can be factored as follows:

$$A = U\Sigma V^*, \qquad (2.230)$$

where U is an $M \times M$ unitary matrix, V is an $N \times N$ unitary matrix, and Σ is an $M \times N$ matrix with nonnegative real values $\{\sigma_k\}_{k=0}^{\min(M,N)-1}$ called *singular values* on the main diagonal and zeros elsewhere. The columns of U are called *left singular vectors* and the columns of V are called *right singular vectors*. As for eigenvalues, $\sigma_{\max}(A)$ denotes the largest singular value and $\sigma_{\min}(A)$ the smallest singular value. Also as for eigenvalues, it is conventional to list singular values in nonincreasing order, $\sigma_{\max}(A) = \sigma_0(A) \geq \sigma_1(A) \geq \cdots \geq \sigma_{N-1}(A) = \sigma_{\min}(A)$. The number of nonzero singular values is the rank of A. The pseudoinverse of A is

$$A^\dagger = V\Sigma^\dagger U^*, \qquad (2.231)$$

where $\Sigma^\dagger$ is the $N \times M$ matrix with $1/\sigma_k$ in the (k,k) position for each nonzero singular value and zeros elsewhere.

Matrices: eigenvalues – singular values (III)

The singular value decomposition and the eigendecomposition are related as follows:

$$
\begin{aligned}
AA^* &= (U\Sigma V^*)(V\Sigma^* U^*) = U\Sigma^2 U^*, \\
A^*A &= (V\Sigma^* U^*)(U\Sigma V^*) = V\Sigma^2 V^*,
\end{aligned}
$$

so the squares of the singular values of A are the nonzero eigenvalues of AA^* and A^*A; that is,

$$\sigma^2(A) = \lambda(AA^*) = \lambda(A^*A), \qquad \text{for } \lambda \neq 0. \tag{2.232}$$

Some special matrices (I)

Unitary and orthogonal matrices A square matrix U is called *unitary* when it satisfies

$$U^*U = UU^* = I. \tag{2.236}$$

Its inverse U^{-1} equals its Hermitian transpose U^*. A real unitary matrix satisfies

$$U^\top U = UU^\top = I, \tag{2.237}$$

and is called *orthogonal.*

Unitary matrices preserve norms for all complex vectors,

$$\|Ux\| = \|x\|,$$

and more generally preserve inner products,

$$\langle Ux, Uy\rangle = \langle x, y\rangle.$$

Some special matrices (II)

Hermitian, symmetric, and normal matrices A *Hermitian* matrix is equal to its adjoint,

$$A = A^*. \tag{2.239a}$$

Such a matrix must be square and is also called *self-adjoint*. A real Hermitian matrix is equal to its transpose,

$$A = A^\top, \tag{2.239b}$$

and is called *symmetric*.

A matrix A is called *normal* when it satisfies $A^*A = AA^*$; in words instead of symbols, it commutes with its Hermitian transpose. Hermitian matrices are obviously normal. A matrix is normal if and only if it can be unitarily diagonalized.

Some special matrices (III)

Circulant matrices A (right) *circulant* matrix is a matrix where each row is obtained by a (right) circular shift of the previous row,

$$C = \begin{bmatrix} c_0 & c_{N-1} & \cdots & c_1 \\ c_1 & c_0 & \cdots & c_2 \\ \vdots & \vdots & \ddots & \vdots \\ c_{N-1} & c_{N-2} & \cdots & c_0 \end{bmatrix}. \tag{2.245}$$

A circulant matrix is diagonalized by the DFT matrix (3.164), as we will see in (3.181b). This means that the columns of the DFT matrix are the eigenvectors of this circulant matrix, and, since the DFT matrix is unitary, these eigenvectors are orthonormal.

Some special matrices (IV)

Toeplitz matrices A *Toeplitz* T matrix is a matrix whose entry T_{ki} depends only on the value of $k-i$. A Toeplitz matrix is thus constant along diagonals,

$$T = \begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{N-1} \\ t_{-1} & t_0 & t_1 & \cdots & t_{N-2} \\ t_{-2} & t_{-1} & t_0 & \cdots & t_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{-N+1} & t_{-N+2} & t_{-N+3} & \cdots & t_0 \end{bmatrix}. \tag{2.246}$$

A matrix in which blocks follow the form above is called a *block Toeplitz* matrix.

Some special matrices (V)

Band matrices A *band* or *banded* matrix is a square matrix with nonzero entries only in a band around the main diagonal. The band need not be symmetric; there might be N_r occupied diagonals on the right side and N_ℓ on the left side. For example, a 5×5 matrix with $N_r = 2$ and $N_\ell = 1$ is of the following form:

$$B = \begin{bmatrix} b_{00} & b_{01} & b_{02} & 0 & 0 \\ b_{10} & b_{11} & b_{12} & b_{13} & 0 \\ 0 & b_{21} & b_{22} & b_{23} & b_{24} \\ 0 & 0 & b_{32} & b_{33} & b_{34} \\ 0 & 0 & 0 & b_{43} & b_{44} \end{bmatrix}. \tag{2.247}$$

Many sets of special matrices are subsets of the band matrices. For example, diagonal matrices have $N_r = N_\ell = 0$, tridiagonal matrices have $N_r = N_\ell = 1$, upper-triangular matrices have $N_\ell = 0$, and lower-triangular matrices have $N_r = 0$. Square matrices have a well-defined *main antidiagonal* running from the lower-left corner to the upper-right corner. An *antidiagonal matrix* has nonzero entries only in the main antidiagonal. A useful matrix is the *unit antidiagonal matrix*, which has ones on the main antidiagonal.

Some special matrices (VI)

Vandermonde matrices A *Vandermonde matrix* is a matrix of the form

$$V = \begin{bmatrix} 1 & \alpha_0 & \alpha_0^2 & \cdots & \alpha_0^{N-1} \\ 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{M-1} & \alpha_{M-1}^2 & \cdots & \alpha_{M-1}^{N-1} \end{bmatrix}. \tag{2.248}$$

When $M = N$, the determinant of the matrix is

$$\det V = \prod_{0 \le i < j \le N-1} (\alpha_i - \alpha_j). \tag{2.249}$$

Many useful concepts in sequence processing use Vandermonde matrices, such as the DFT matrix.